Comparing the results of Extended Gumbel Type-2 Distributions found in literature with those obtained by the free CumFreq program

R.J. Oosterbaan
September 2022   website: www.waterlog.info

Abstract
A. A. Ogunde et al. have given data to which they have fitted different upgraded Gumbel probability distributions like the Extended Gumbel Type-2 Distribution.. In this paper these data are used in the free CumFreq program with the option to find the best fitting of all distribution dealt with in this software and compare the results. One reason for this operation is to demonstrate that CumFreq is easy to operate and all computations are done automatically. Another reason is to try to find even better fitting distributions than those presented by the authors mentioned before while the number of parameters is less, which enhances the robustness.

Specifications of references in abstract
1. A. A. Ogunde et al. (2022)
   Extended Gumbel Type-2 Distribution: Properties and Applications
2. R.J.Oosterbaan, CumFreq
   https://www.waterlog.info/cumfreq.htm

Contents
1. Brief introduction to CumFreq
2. Survival times of guinea pigs infected with virulent tubercle bacilli
3. Remission times of bladder cancer patients
4. Conclusion
5. References
6. Appendices
   A1. CumFreq input menu
   A2. Selection options in CumFreq

1. Brief introduction to CumFreq

After entering the data in the input menu (for example by copying from excel and pasting, see Appendix A1)), one can select a preferred probability distribution (see Appendix A2) or choose "*best of all*". Clicking on "Save-Run" CumFreq will then perform the probability distribution fitting automatically, show the results, and provide the opportunity to see graphs.

CumFreq uses plotting positions: $F_c = r / (N+1)$ where r is the rank number of the data arranged in ascending order: $r = 1, 2. 3, \ldots . N$ with N being the total number of data to. Here F_c is the observe cumulative frequency or probability (Makkonen 2006). With the help of F_c it is fairly simple to find the parameters of a large number of well known probability distributions.

For more detail consult R.J. Oosterbaan (2022) How to derive a probability distribution from a data set using the simple method of plotting positions and the free CumFreq model

2. Survival times of guinea pigs infected with virulent tubercle bacilli.

According to CumFreq the best of all distributions for the survival times of guinea pigs is the Fréchet (Fischer-Tippett Type II) cumulative probability distribution function (CDF or Fc):

$$CDF = Fc = \exp \left[- \left\{ (X\text{-}C) / \exp (\text{–}B/A) \right\}^A \right]$$

The Fc values can be found by the plotting positions: Fc = r / (N+1) where r is the rank number of the data arranged in ascending order: r = 1, 2. 3, ... N with N being the total number of data (Makkonen 2006, Oosterbaan 2022).

Using the transformed values Ft = ln { –ln (Fc) } and Xt = ln (X–C) one can write:
Ft = A*Xt + B , where A and B can be found from a simple linear regression of Ft upon Xt, while the term C is to be numerically optimized. The distribution has three parameters.

In the case of survival times of guinea pigs, CumFreq gave the following result:
A = –4.70E+000   B = 5.48E+000   C = –1.97E+000

The histogram together with the probability density function is shown in the figure 1 hereunder.
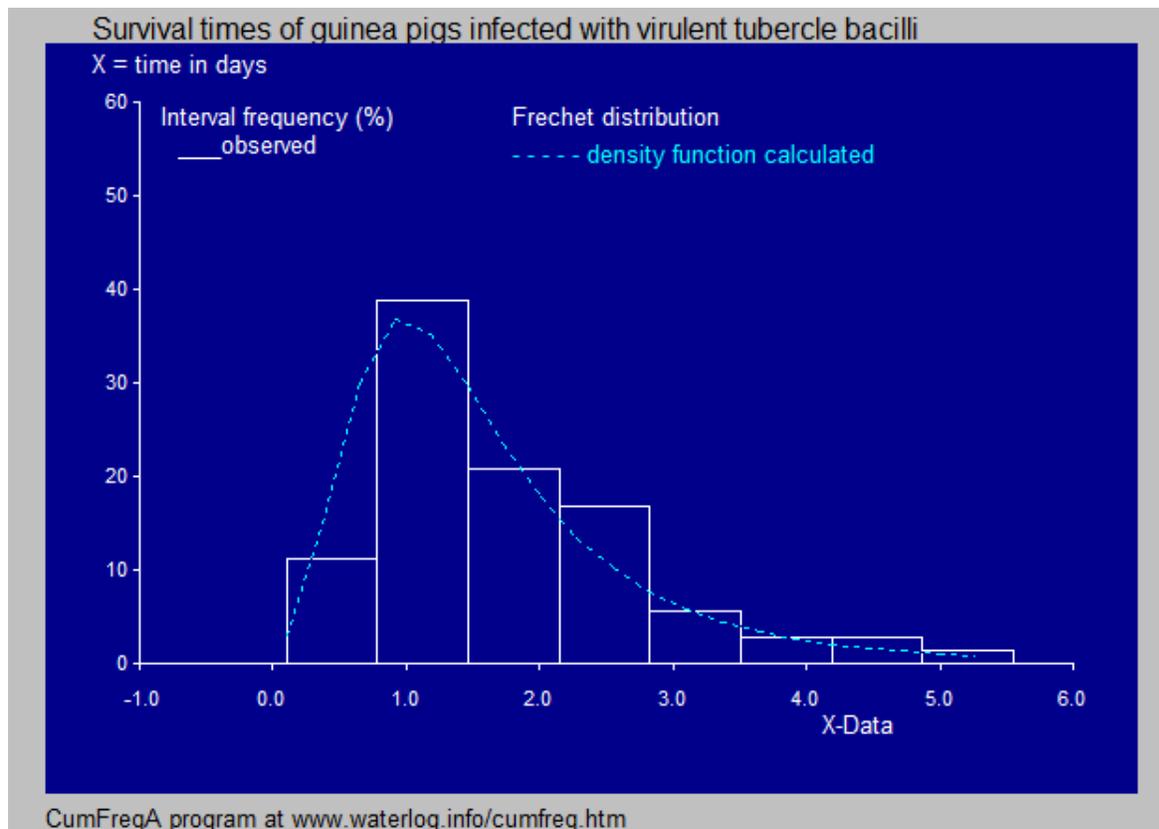

Figure 1.

A. A. Ogunde et al, using the Extended Gumbel Type-2 Distribution (EGTT), working with four parameters, have reached the following result:
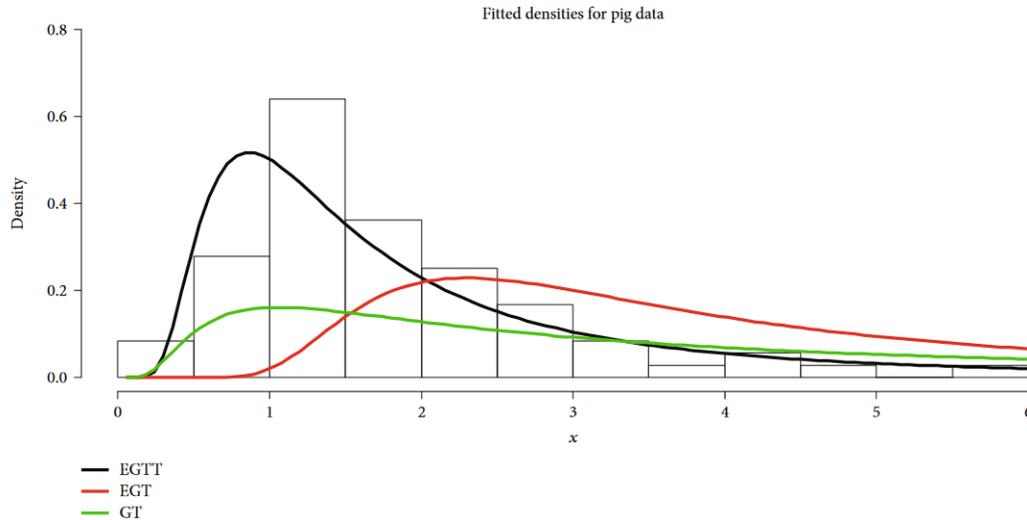
Fitted densities for pig data

FIGURE 6: Fitted densities for pig data.

Figure 2.

The density curve in figure 1 fits better to the observed data represented by the histogram then the one marked EGTT (black curve) in figure 2. Hence, in this case, the Fréchet distribution (with 3 parameters) deserves preference over the Extended Gumbel Type-2 Distribution even though it uses 4 parameters.

The option in CumFreq to detect the best fitting of all distributions results in a ranking according to goodness of fit. The next table gives the first eight CDF's for remission times of bladder cancer patients.

Distribution name *)	Average of absolute differences between calculated and observed cumulative frequencies #)
1. Frechet (Fisher-Tippet Type 2)	1.91 %
2. Gumbel generalized	1.96 %
3. Log-normal optimized	2.06 %
4. Burr generalized	2.10 %
5. Generalized extreme value (GEV)	2.34 %
6 Log-normal standard	2.37 %
7. Laplace generalized	2.49 %
8. Log-logistic	2.50 %
9 Laplace standard	2.85 %

*) Generalization means raising X to the power P.
   Optimization means adjusting mean and standard deviation
#) 0–1% excellent, 1–2% very good, 2–3% good, 3 – 4% medium, > 4 % poor

A. A. Ogunde et al. use different indexes for the goodness of fit so that here a comparison cannot be made.

3. Remission times bladder cancer patients

According to CumFreq the best of all distributions for the remission times of bladder cancer patients is the Burr (generalized Pareto-Lomax) cumulative probability distribution function (CDF or Fc):

$$\text{CDF} = \text{Fc} = 1 - [\, B / (X^A + B]^E$$

The generalization is done by adding the exponent A.

The Fc values can be found by the plotting positions: Fc = r / (N+1) where r is the rank number of the data arranged in ascending order: r = 1, 2. 3, ... N with N being the total number of data (Makkonen 2006, Oosterbaan 2022).

Using the transformed values $Ft = \ln(1 - Fc)$ and $Xt = \ln\{\, B /(X^A + B)$ one can write: $Ft = E*Xt$, where A can be found from a simple ratio method while A and B need to be determined numerically. The distribution has three parameters.

In the case of remission times of bladder cancer patients, CumFreq gave the following result:
A = 1.63E+000   B = 2.97E+001   E = 1.33E+000

The histogram together with the probability density function is shown in the figure 3 hereunder.
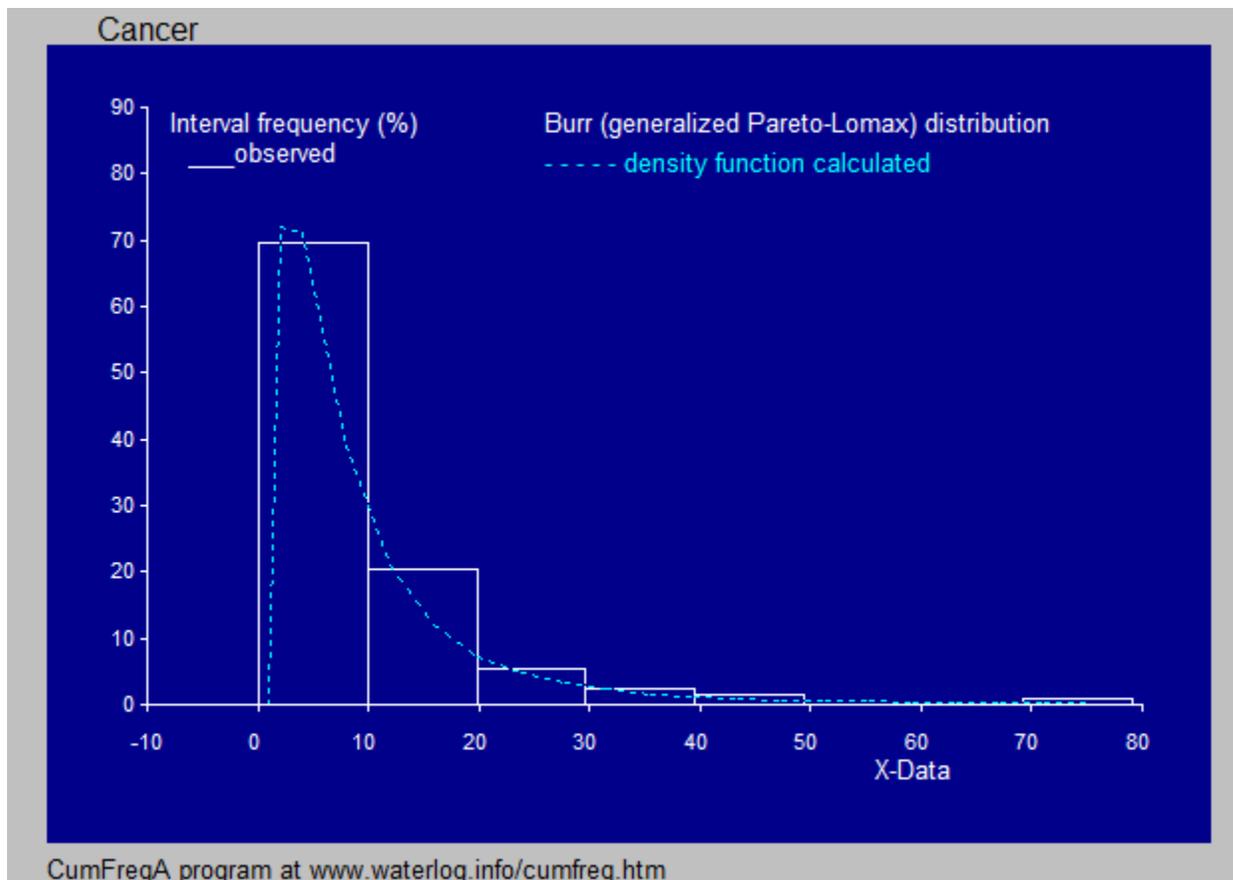
Figure 3.

A. A. Ogunde et al, using the Extended Gumbel Type-2 Distribution EGTT), working with four parameters, have reached the following result:
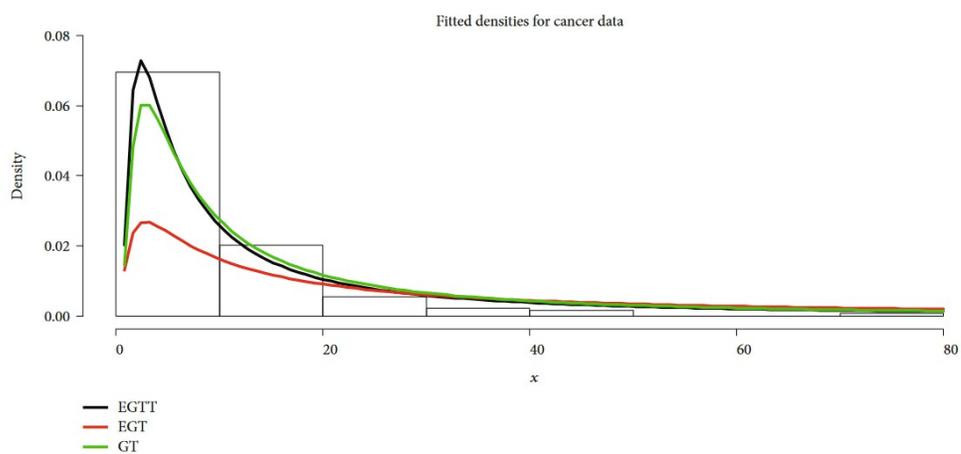

FIGURE 8: Fitted densities of the cancer data.

Figure 4.

The density curve in figure 3 fits to the observed data represented by the histogram as well as the one marked EGTT (black curve) in figure 4. Hence in this case the Burr distribution (with 2 parameters) gives similar results as the Extended Gumbel Type-2 Distribution with 4 parameters. However, as the Burr distribution uses a smaller number of parameters, it deserves preference.

The option in CumFreq to detect the best fitting of all distributions results in a ranking according to goodness of fit. The next table gives the first eight CDF's for remission times of bladder cancer patients.

Distribution name *)	Average of absolute differences between calculated and observed cumulative frequencies #)
1. Burr generalized	0.75 %
2. Generalized extreme value (GEV)	0.78 %
3. Fisk distribution	1.06 %
4. Log-normal optimized	1.12 %
5. Dagum generalized	1.23 %
6. Gumbel generalized	1.40 %
7. Laplace generalized	1.60 %
8. Rayleigh generalized	1.65 %

*) Generalization means raising X to the power P.
   Optimization means adjusting mean and standard deviation
#) 0–1% excellent, 1–2% very good, 2–3% good, 3 – 4% medium, > 4 % poor

In the table, the first two distributions give excellent results while the third and fourth do not fall much behind.

A. A. Ogunde et al. use different indexes for the goodness of fit so that here a comparison cannot be made.

4. Conclusion

For practical purposes CumFreq is easy to use, versatile, and recommendable (Oosterbaan, 2019).

5. References

L. Makkonen, 2006. Plotting Positions in Extreme Value Analysis**.** Journal of applied meteorology an climatology, 01 Feb 2006, pp. 334-340.. https://doi.org/10.1175/JAM2349.1
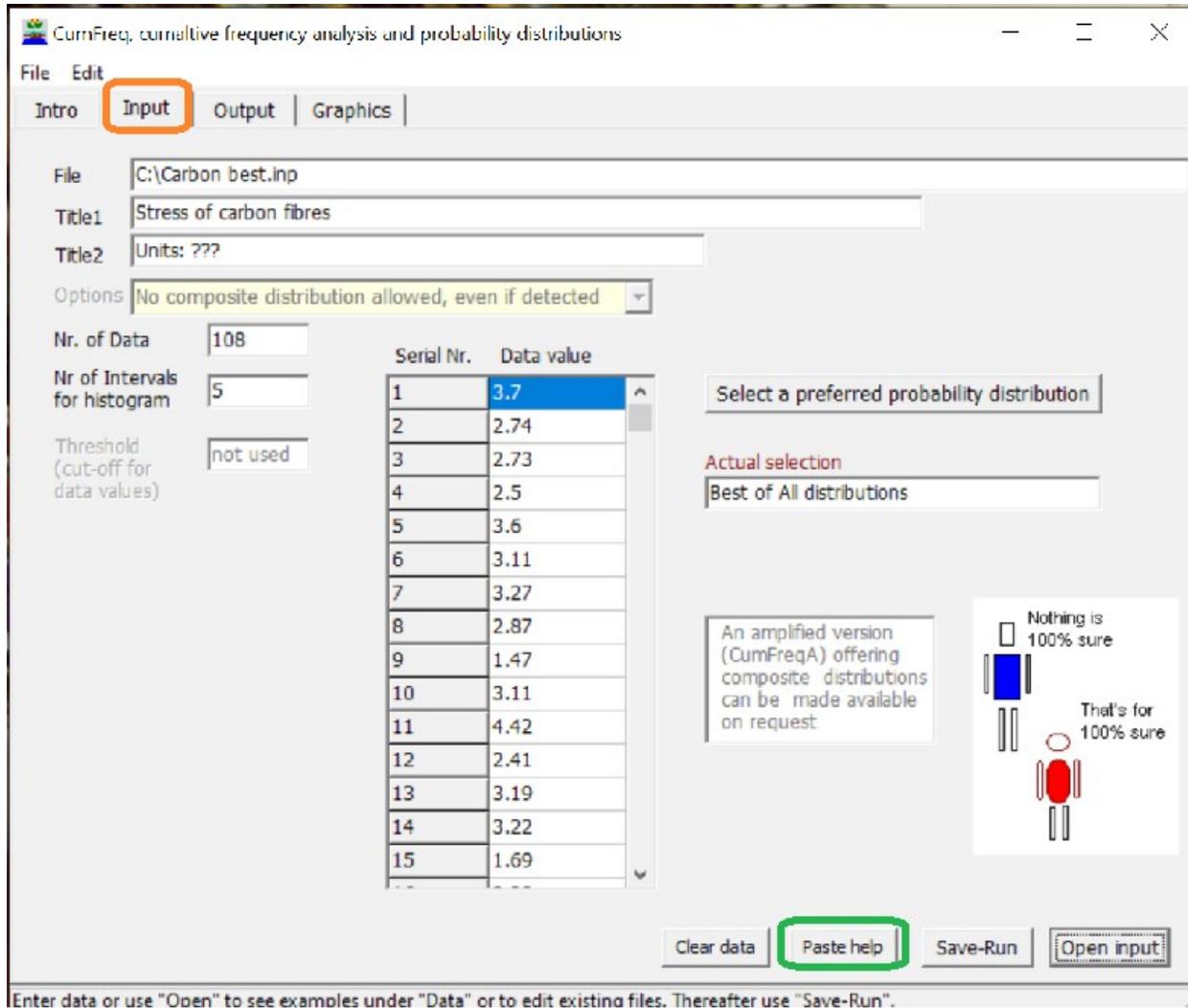
R.J.Oosterbaan, CumFreq software: https://www.waterlog.info/cumfreq.htm

R.J. Oosterbaan (2022), CumFreq, plotting positions:
How to derive a probability distribution from a data set using the simple method of plotting positions and the free CumFreq model

R.J. Oosterbaan, 2019. Software for generalized and composite probability distributions. International Journal of Mathematical and Computational Methods, Volume 4, 2019, 1-9.

9. Appendices

Appendix A1. CumFreq input menu

Appendix A2. Selection options in CumFreq

CumFreq select distributions — □ ✕

Make a selection then click "Confirm"

- ● Best of All Distributions
- ○ Normal (standard)
- ○ Normal (optimized)
- ○ Log-normal (standard)
- ○ Log-normal (optimized)
- ○ Root-normal
- ○ Square-normal
- ○ Generalized normal
- ○ Logistic (standard)
- ○ Log-logistic
- ○ Logistic generalized
- ○ Burr distribution
- ○ Cauchy (standard)
- ○ Cauchy generalized
- ○ Dagum distribution
- ○ Exponential (standard)
- ○ Exponential generalized
- ○ Exponential general. mirr.
- ○ Gamma (Erlang)

- ○ Fisk (adjusted)
- ○ Fisher-Tippet III
- ○ F-T III mirrored
- ○ Frechet (typical)
- ○ Frechet mirrored
- ○ GEV (Gen. Extreme. Value)
- ○ Gompertz (generalized)
- ○ Gumbel (standard)
- ○ Gumbel generalized
- ○ Gumbel mirrored
- ○ Gumbel mirrored generalized
- ○ Kumaraswamy generalized
- ○ Laplace (standard)
- ○ Laplace generalized
- ○ Rayleigh generalized
- ○ Student (1 or 2 d.f.)
- ○ Weibull (standard)
- ○ Weibull mirrored

Confirm